Anal Bioanal Chem (2011) 399:387–401
DOI 10.1007/s00216-010-4342-z

ORIGINAL PAPER

Metabolic fingerprinting as a tool to monitor whole-cell biotransformations

Catherine L. Winder · Robert Cornmell · Stephanie Schuler · Roger M. Jarvis · Gill M. Stephens · Royston Goodacre

Received: 16 July 2010 / Revised: 12 October 2010 / Accepted: 13 October 2010 / Published online: 31 October 2010

Abstract Fourier transform infrared (FT-IR) spectroscopy was employed as a rapid high-throughput phenotypic typing technique to generate metabolic fingerprints of *Escherichia coli* MG1655 pDTG601A growing in fed-batch culture, during the dioxygenase-catalysed biotransformation of toluene to toluene *cis*-glycol. With toluene fed as a vapour, the final toluene *cis*-glycol concentration was 83 mM, whereas the product concentration was only 22 mM when the culture was supplied with liquid toluene. Multivariate statistical analysis employing cluster analysis was used to analyse the dynamic changes in the data. The analysis revealed distinct trends and trajectories in cluster ordination space, illustrating phenotypic changes related to differences in the growth and product formation of the cultures. In addition, partial least squares regression was used to correlate the FT-IR metabolic fingerprints with the levels of toluene *cis*-glycol and acetate, the latter being an indicator of metabolic stress. We propose that this high-throughput metabolic fingerprinting approach is an ideal tool to assess temporal biochemical dynamics in complex biological processes, as demonstrated by this redox biotransformation. Moreover, this approach can also give useful information on product yields and fermentation health indicators directly from the fermentation broth without the need for lengthy chromatographic analysis of the products.

Keywords Fourier transform infrared spectroscopy · *Escherichia coli* · Biocatalysis · Metabolomics · Toluene *cis*-glycol · Fed-batch culture

Electronic supplementary material The online version of this article (doi:10.1007/s00216-010-4342-z) contains supplementary material, which is available to authorized users.

C. L. Winder (✉) · R. M. Jarvis · R. Goodacre
School of Chemistry, The University of Manchester,
131 Princess Street,
Manchester M1 7DN, UK
e-mail: C.Winder@manchester.ac.uk

R. Cornmell · S. Schuler · G. M. Stephens
School of Chemical Engineering and Analytical Sciences,
Manchester Interdisciplinary Biocentre,
The University of Manchester,
131 Princess Street,
Manchester M1 7DN, UK

G. M. Stephens
Department of Chemical and Environmental Engineering, Process and Environmental Research Division, University of Nottingham,
University Park,
Nottingham NG7 2RD, UK

Introduction

Many manufacturing industries are beginning to replace traditional chemical processes with bioprocesses because of their chemical specificity, desirable reaction kinetics and benign conditions. The regio- and stereo-selectivity are frequently unmatched by chemical routes, therefore providing an attractive ‘green route’ to manufacture chemical products that cannot be produced in alternative reactions. Redox biotransformations should provide many new routes to the manufacture of both chiral and non-chiral speciality chemicals. However, the majority of redox reactions require the use of whole cells as biocatalysts to provide a means of cofactor regeneration [1]. The problem with using whole cells as biocatalysts is that the product yields are limited by substrate and/or product toxicity. To reduce the effects of toxic substrates on whole-cell biotransformation, specific growth conditions such as two-liquid phase systems [2] and

gradual feeding procedures may be employed. However, the key to improving whole-cell biotransformations is a clear understanding of the limitations at all levels of the bioprocess. To understand the effect of the bioprocess on the physiology of the cultures, we have employed a metabolic fingerprinting technique to profile the cultures at a cellular level.

In this study, we used recombinant *Escherichia coli* expressing toluene dioxygenase to produce toluene *cis*-glycol (TCG) from toluene (Fig. 1) during growth in glucose-limited fed-batch culture [3]. The use of whole cells in redox biotransformations produces improved yields and is a "greener" approach than using chemical methods. The resulting synthetically important products are useful as precursors in new or alternative routes in chiral synthesis [4, 5]. Organic solvents such as toluene are extremely toxic to the cells in the region of 0.1% (*v*/*v*) [6], and the build-up of substrate results in an environment detrimental to the viability of the culture and product yield. Fourier transform infrared (FT-IR) spectroscopy was used as a phenotypic typing technique which is routinely used to generate metabolic fingerprints of microbial cultures [7–10]. When coupled to multivariate statistical analysis, FT-IR spectroscopy has been used in numerous studies, including the differentiation of bacteria to sub-species levels [7, 11–14], functional genomics screens in *E. coli* [15], the effects of pharmaceuticals on microbial cultures [16] and measuring the concentrations of target molecules in microbial fermentations [17–19]. The sensitivity of this approach has also been used in biotechnological processes to model dynamic fluctuations in silage fermentations [10], as a tool in bioprocess monitoring in acetone–butanol fermentations by *Clostridium* strains [20], and to determine the effect of temperature on the membrane order and fluidity in Gram-negative bacteria [21]. FT-IR spectroscopy presents itself as an ideal high-throughput screening technique due to the speed in which the fingerprint may be generated (typically 10 secs to 1 min per sample), minimal sample preparation, automation (96- or 384-well plates) and relatively low expense.

To our knowledge, FT-IR has not been previously employed to monitor metabolic fingerprints in *E. coli* biotransformations. In this paper, we demonstrate that, in combination with appropriate multivariate analyses (chemometrics), FT-IR spectroscopy has sufficient sensitivity to assess the biochemical fluctuations during toluene *cis*-glycol production. This high-throughput method is an ideal tool, since the analysis of large sample numbers is required to fully assess the time course of the bioprocess. We have applied two approaches to add the substrate (toluene) to the cultures, either as a vapour or liquid feed. Classical microbiological measurements were taken to assess growth and cell physiology, in addition to the phenotypic analysis by FT-IR metabolic fingerprinting of the cell composition. Cluster analysis (PC-CVA) and partial least squares regression (PLSR) were used to assess the temporal dynamics of the fermentation and to simultaneously determine the effect of the product (toluene *cis*-glycol) and acetate in the exometabolome on the fingerprint of the cultures.

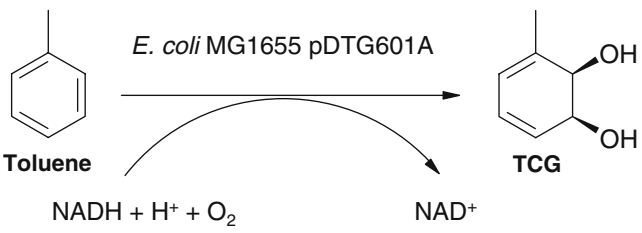

Fig. 1 Enzymatic conversion of toluene to toluene *cis*-glycol by toluene dioxygenase

Experimental

Growth conditions

E. coli MG1655 [22] was transformed [23] with pDTG601A (kindly supplied by Rebecca Parales, University of California, Davis). This plasmid carries the toluene dioxygenase gene under the *tac* promoter [24], and this host-expression system was used as a biocatalyst to convert toluene to TCG (Fig. 1). The conditions for growth and biotransformation had previously been developed using *E. coli* JM109 as described by Phumathon and Stephens [3], so we tested both host strains for the biotransformation to check that the conditions were suitable for MG1655. Preliminary experiments (data not shown) showed that there was no significant difference in the TCG yields between the original JM109 host and MG1655, indicating that the biotransformations could be operated as previously described. This was an important finding since the whole genome sequence of MG1655 is known and so will advance the understanding of the effect of the biotransformation on the host in future studies. Biotransformations were performed in a Pierre Guerin bioreactor (Mauze, France) fitted with a 2-L vessel and operated at an initial working volume of 1.25 L in double-strength MS medium [25, 26] with 100 μg mL^{-1} ampicillin. The vessel was inoculated with *E. coli* MG1655 cultured in MSX medium [26] with 100 μg mL^{-1} ampicillin for 18 h at 37 °C to an optical density$_{680\ nm}$ of 1.0. The culture was grown at 37 °C, at pH 6.8±0.1 (controlled by automatic addition of 1 M NaOH), with an air flow of 30 L h^{-1}. The dissolved oxygen concentration was maintained above 5%, by automatic control of the stirrer speed (minimum 200 rpm). Foaming

was controlled by the automatic addition of polypropylene glycol P2000 (3% *v*/*v*) when required.

The biomass of the cultures was measured by monitoring the optical density at 680 nm and dry weight of the cultures. When the culture reached a dry weight of 0.5 g L^{-1}, the pH was increased to 7.2 to prevent decomposition of TCG. The culture was allowed to adapt to the pH change for 30 min, and then isopropyl-β-D-thiogalactoside (IPTG; 1 mM) was added to induce expression of toluene dioxygenase, and the toluene feed was started (if applicable). Toluene vapour was added by sparging air (15 L h^{-1}) via a glass sinter through a reservoir of liquid toluene (300 mL in a 500-mL Dreschel bottle) and mixing this with pure air (15 L h^{-1}) before passing the mixture into the culture via a sterile filter. Liquid toluene (*liquid feed*) was supplied using a syringe pump (model 100 series, KD Scientific, USA) at an initial rate of 0.1 mL h^{-1} for 2 h. The flow rate was increased at a rate of 0.1 mL h^{-1} to a final rate of 0.5 mL h^{-1} and maintained for a further 19 h. The toluene was added to the vessel through a glass Pasteur pipette to reduce the droplet size and ensure that the droplets left the orifice rapidly. This ensured efficient mixing and also minimised exposure of the toluene droplets to the air flow to reduce evaporative loss. Control cultures were grown without toluene, according to the same general procedure. In all cases, a glucose salt solution [25, 26] was fed to the culture after the consumption of the glucose in the initial culture medium. The initial feed rate was 8 mL h^{-1}, and the rate was increased whenever the oxygen demand decreased, or offline measurement of glucose concentration indicated that the culture was glucose-limited. The feed rate was adjusted to continue the exponential growth of the bacteria and also to enable cofactor recycling for the biotransformation. The maximum feed rate was 12 mL h^{-1}.

In total, 13 fed-batch cultures were grown and analysed (Table 1). These included four controls where no toluene was added, five cultures where toluene was introduced in vapour form and four cultures using liquid-fed toluene.

Analytical methods

Biomass concentrations were determined by monitoring the optical density (680 nm) and the dry weight of the culture. The dry weight of the cultures was measured by filtering a known volume of culture through a pre-weighed 0.2-μm membrane filter (Whatman, UK). The filters were dried until consistent weights were obtained, and the dry weight of the cultures was calculated. Viable counts and culture purity were measured by serial dilution in 0.9% saline solution, and the preparation of replicate spread plates on LB agar. The viability of the cultures was consistent between the experimental conditions (data not shown). Samples of culture supernatant were prepared by centrifu-

Table 1 Growth and product formation in the replicate fed-batch cultures with different toluene feeding regimes

Repeat	No toluene added						Vapour-fed cultures						Liquid-fed cultures					
	1	2	3	4	Av	SD	1	2	3	4	5	Av	SD	1	2	3	4	Av	SD
Max biomass ($g_{dry\ weight}$ L^{-1})	7.22	4.40	6.96	6.73	6.33	1.30	6.42	7.42	6.69	7.82	9.50	7.57	1.22	7.29	6.39	6.64	5.43	6.44	0.77
Specific biomass yield[a]	0.18	0.13	0.15	0.20	0.17	0.03	0.16	0.16	0.15	0.18	0.25	0.18	0.04	0.23	0.18	0.15	0.12	0.17	0.05
Acetate/glucose ratio[b]	0.26	0.29	0.22	0.25	0.25	0.03	nd	nd	0.15	0.25	0.15	0.18	0.06	0.33	0.30	0.24	0.24	0.28	0.05
Maximum TCG concentration (mM)	n/a	n/a	n/a	n/a	n/a	n/a	91.02	86.90	73.03	80.30	58.20	77.89	12.95	26.38	19.68	9.60	0.28	13.99	11.45
Specific TCG[c]	n/a	n/a	n/a	n/a	n/a	n/a	1.79	1.48	1.38	1.30	0.77	1.34	0.37	0.46	0.39	0.18	0.01	0.26	0.20

Av average, *SD* standard deviation, *n/a* not applicable, *nd* not done

[a] Grams biomass per gram glucose

[b] Grams acetate produced per gram consumed glucose

[c] Millimolar TCG produced per gram biomass produced

gation for 3 min at 11,500×*g* and analysed for residual glucose, acetate and TCG. The glucose concentration was determined using a blood glucose metre (Medisense Precision QID, UK). TCG concentrations were determined by filtering the samples through a 0.2-μm syringe filter, storing at −20 °C and then analysing each sample in triplicate [26]. Acetate concentrations were measured using a Hewlett Packard 5890 Series II gas chromatograph equipped with a 7673 auto-injector and flame ionisation detector. Samples of culture supernatant (1 mL) were acidified by adding 100 μL of 15% (*v*/*v*) HCl, and aliquots (1 μL) were injected in triplicate onto a PoraPLOT Q column (Chrompack), with the injector and detector temperatures at 200 °C. The temperature programme started with an initial temperature of 130 °C for 1 min and was increased to 185 °C at a rate of 10 °C min^{-1} and maintained for 1 min. The concentration of acetate was calculated by correlation with authentic sodium acetate (Sigma-Aldrich) standards.

Metabolic fingerprinting was performed by FT-IR analysis; in this investigation, we refer to the 'metabolic fingerprint' or 'fingerprint' as the region on the spectra in the range of 4,000–900 cm^{-1}. Duplicate samples were taken from each culture for the FT-IR analysis. The culture was centrifuged for 3 min at 11,500×*g* to recover the cells and washed three times with sterile physiological saline (0.9% NaCl) before resuspension in saline solution. The OD was measured, and samples were adjusted to the same biomass concentration by dilution with saline solution. Prior to use, a 96-well zinc selenide plate was rinsed with isopropanol and deionised water and allowed to dry at room temperature. The combination of both biological and analytical replicates was employed in the FT-IR analysis to measure the biological variance in the data. The employment of 'analytical or machine replicates' [7] may be averaged to reduce the heterogeneity of the biological replicates in the analysis of the data [27]. Replicate aliquots (n=4, 20 μL) of the biological replicates were evenly applied onto the plate and dried at 50 °C for 15 min. The ZnSe plate was loaded onto a motorised microplate module HTS-XT™ attached to an Equinox 55 spectrometer (Bruker Optics Ltd., UK). The motorised module of this instrument introduces the plate into the airtight optics of the instrument, in which tubes of desiccant are contained to remove moisture [7, 28]. A deuterated triglycine sulphate detector was employed for transmission measurements of the samples to be acquired. An empty well was used to take a reference background measurement. Spectra were collected over the wavenumber range of 4,000 to 600 cm^{-1} under the control of a computer programmed with Opus 4, operated under MS Windows 2000. Spectra were acquired at a resolution of 4 cm^{-1}, and 64 spectra were co-added and averaged to improve the signal-to-noise ratio as previously described [7]. The collection time for each spectrum was approximately 1 min, and the spectra are usually displayed in terms of absorbance (*vide infra*). The data were converted to ASCII format by an in-house conversion program.

Multivariate data analysis of the FT-IR spectra

Preprocessing The ASCII data were imported into Matlab version 6 (The MathWorks, Inc., MA, USA) [29], and all of the statistical analyses were performed in Matlab, and the scripts are available on the website (www.biospec.net). To minimise problems arising from baseline shifts, Matlab was used to correct for CO_2 vibrations; when present, the CO_2 peaks at 2,403–2,272 cm^{-1} and 683–656 cm^{-1} were removed and filled with a trend to maintain a baseline representative of the spectra. An extended multiplicative signal correction preprocessing method was applied to the entire data set [30], and the spectra were normalised according to zero mean and unit variance (auto-scaling). The statistical analysis of the data was performed on the spectral region between 4,000 and 900 cm^{-1}.

Cluster analyses The unsupervised data reduction method, principal components analysis (PCA; [31]), was performed on the spectra to reduce the dimensionality of the multivariate data whilst preserving the variance, prior to canonical variate analysis (CVA). CVA is a supervised technique which discriminates between groups on the basis of the retained principal components with *a priori* knowledge of the analytical replicates [32, 33]. CVA was programmed to minimise 'within-group' variance and maximise 'between-group' variance, and as this process uses information based on the biological replicates from each bioreactor growth run, it does not bias the analysis and allows any natural trends or time-dependent trajectories to be observed. The statistical significance is displayed on the scores plots to indicate the 95% χ^2 distribution with 2 degrees of freedom [34].

Partial least squares regression Models were generated from the independent measurements of both acetate and TCG concentrations detected in the exometabolome. PLSR [30, 35] was used, as described previously [36–38]. PLSR is a supervised multivariate method that applies a linear fixed regressor model of the form:

$$Y = AX + B$$

The algorithm finds the fundamental rela tionships between the predictor (Y) variables and response (X) variables in the form of latent variables (A), with the error (B). Each set of PLSR predictions were constructed using a bootstrapping approach [39–41], whereby a Monte Carlo simulation was

used to iteratively ($n=10^2$) sub-sample training and cross-validation spectral sub-sets. The data set was split into training and cross-validation sub-sets according to the biological replicates. In general, four biological replicates were performed for each time point, and in this instance, the data were split 1:1 to provide sub-sets for the training and cross-validation samples. If an odd number of biological replicates was present in the data (such as in the vapour biotransforming cells), the data were split 2:1 for the training and cross-validation. PLSR models were then evaluated using these randomly selected training and test samples using full cross-validation [37, 42]. With cross-validation, the model is trained on one set of data (the training set) and the optimal number of latent variables selected by projection of the second set of hold-out data (the cross-validation set), to determine the number of dimensions required to obtain the *minimum* root mean squared error (RMSE) of cross-validation. Finally, the 95% bootstrap confidence limits were calculated using the residual error from the least squares fit on the 1,000 training set predictions, and a simple count was performed to determine the number of training and cross-validation predictions within these limits. Given the large number of bootstrap iterations performed for the analysis, every sample was selected more than once in both the training and cross-validation sub-sets. Latent variables for each bootstrap model were automatically selected based upon the minimum of the root mean squared error for the cross-validation predictions, and therefore a range of values was used across the 1,000 models. The number of latent variables representing the average minimum RMSE across all validation predictions is used as a basis from which to summarise the results in Table 2.

Results

Growth of *E. coli* in fed-batch cultures

A fed-batch culture system was employed in this study [3] to maximise the biomass and product yield and provide a realistic comparison with large-scale bioprocesses. Initially, control cultures were grown without toluene, but with induction of the dioxygenase enzyme. Exponentially growing cultures were established, and when the biomass concentration reached 0.5 $g_{dry\ weight}\ L^{-1}$, the pH was adjusted to 7.2, and the dioxygenase enzyme was induced with IPTG. Once the glucose had been consumed, a glucose salt solution was added to maintain exponential growth of the cultures. To assess the effect of the fed-batch cultivation on the metabolic fingerprints, we included four control cultures in the analysis of which the growth profiles and acetate production are shown in Figs. 2a, b. In one of the cultures (repeat 3), the recombinant protein was not induced until the biomass concentration was 1.54 $g_{dry\ weight}\ L^{-1}$, so that we could analyse the effect of late induction on the metabolic fingerprints. Slight differences were observed in the growth properties of the repeated cultures (Table 1), the most notable in the second repeat. The final biomass concentration and specific biomass yield was lower than average in the second culture, and the acetate-to-glucose ratio was higher than average, suggesting a change in the metabolism of the culture.

Analysis of the metabolic fingerprints Variations in the cellular metabolism were investigated by FT-IR metabolic fingerprinting, using cell samples which had been washed thoroughly to remove any extracellular metabolites and

Table 2 Summary of the PLS analysis showing the relationship between the metabolic fingerprints (FT-IR spectra) extracellular TCG and acetate concentration

	Number of latent variables	Independent calibration values		$RMSE_C$[a]	$RMSE_V$[b]
		Number	Range		
Acetate					
Control	8	214	3–170 mM	17.38±4.32	24.58±1.84
Vapour	9	191	4–180 mM	23.64±3.97	31.73±2.51
Liquid	5	198	19–180 mM	20.38±4.56	30.68±4.81
TCG					
Vapour	7	383	0–91 mM	11.18±3.29	10.27±3.06
Liquid	9	198	0–26 mM	3.34±0.43	5.05±0.57

The concentrations of acetate and TCG were determined by GC-FID and LC, respectively, from the samples of the exometabolome. The number of calibration variables and range applied in the models are presented in addition to the average and standard deviation root mean squared error of calibration in millimolar and average and standard deviation root mean squared error of cross-validation in millimolar

[a] Average and standard deviation root mean squared error of calibration in millimolar

[b] Average and standard deviation root mean squared error of cross-validation in millimolar

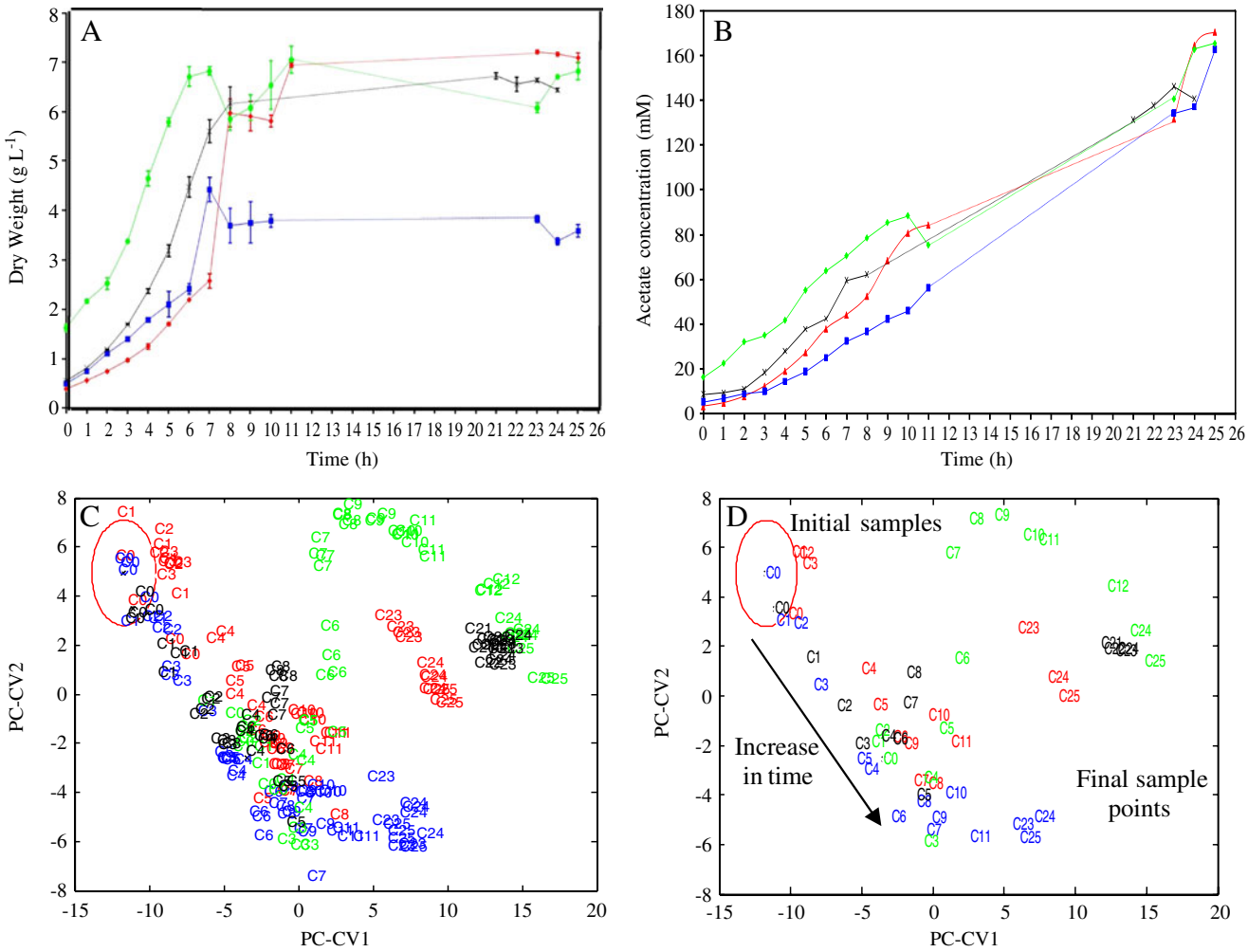

Fig. 2 The growth profile (**a**) of *E. coli* in fed-batch cultures without toluene, with the corresponding extracellular acetate (**b**) concentration. The IPTG was added directly after the 0-h time point. The *error bars* indicate the standard deviation of the analytical replicate data points. PC-CVA performed on the metabolic fingerprints generated by the FT-IR analysis (**c**) PCs 15 (total cumulative variance 99.47%) was used by the CVA algorithm with a priori knowledge of analytical replicate. The variance represented in the canonical variates is 7.5% (CV1) and 2.1% (CV2). The sampling time in hours is referred to as numbers after the letter code. For ease of visualisation, the group mean centres of each replicate are shown in **d**, whilst in **c**, all of the data points are shown and the *circle* indicates the boundary of the 95% confidence region for one particular sample. Each point would have the same area representing the 95% confidence region; however, only one circle is shown for clarity. The *colours* indicate the biological replicates (replicates 1, *red*; 2, *blue*; 3, *green*, 4; *black*); three analytical replicates were used in the measurements of biomass and acetate concentration, and four analytical replicates were used in the FT-IR analysis

medium components. Representative metabolic fingerprints of *E. coli* from the three experimental conditions during fed-batch growth are illustrated in Fig. 3. As expected, these FT-IR spectra (and indeed all others) show broad and complex contours that are typical of microbial samples [7, 29], illustrating that there were few observable qualitative differences between the spectra. Therefore, multivariate statistical analysis was employed to analyse these data. Discrimination between samples at different time points was not observed with unsupervised PCA, although using PCA for data reduction prior to supervised CVA (in which the *a priori* knowledge was based on the analytical replicates) resolved differences and similarities between the spectral fingerprints. The first and second canonical variate (CV) scores were plotted to identify relationships or variations between the samples. A similar trajectory was observed for three of the four repeat cultures, whereby a trend (through CV 2) was observed with increasing culture time (Fig. 2c, d). However, a noticeable shift was observed in the trajectory of the samples from the third culture, in which we had delayed the induction of the recombinant protein until the biomass of the culture was greater than the other cultures (Fig. 2d, illustrated in green).

The samples from the second culture (which exhibited a low biomass yield) were displaced in the CVA space from the end-point samples of the other cultures (illustrated in

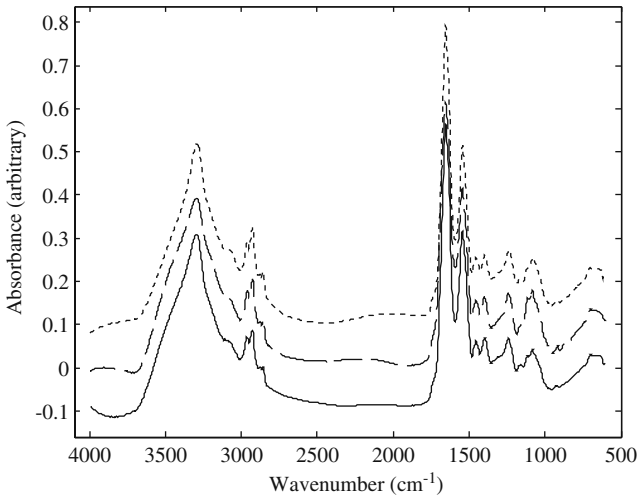

Fig. 3 Typical FT-IR spectra of *E. coli* cultured without toluene (*solid line*), in the presence of vapour toluene (*dashed line*) and liquid toluene (*dotted line*). The spectra representing the cells cultured in the presence of toluene were offset

red, green and black) and clustered in close proximity to the earlier sample points. The CVA (Fig. 2d) indicates that there was less variation in the mid- and end-point samples of the second culture compared with the other culture repeats included in the analyses. The acetate-to-glucose ratio was higher than average in the second repeat compared with the other cultures, and the biomass yield was lower than average. The changes noted in the analysis of the metabolic fingerprints reflect the observed differences in growth efficiency of the cultures and acetate production between the replicate cultures. The loadings of the first PC-CV (Electronic Supplementary Material Figure S1A) which account for 7.5% of the variance in the data show widespread changes in the biological components of the cells, including positive changes in vibrations between 3,000–2,800 and 1,600–1,500 cm^{-1} which are attributed to fatty acid and amide II vibrations, respectively, whilst negative changes were observed in the regions of the loading plot which are attributed to amide I (1,750–1,650 cm^{-1}) and polysaccharides (1,200–900 cm^{-1}). The first PC-CV indicates general differences between the early, mid and late time points of the cultures, whereas the second PC-CV represents the subtle changes between the individual time points, and the loading plot (Electronic Supplementary Material Figure S1A) indicated the predominant changes to be in the polysaccharide region (1,200–900 cm^{-1}). It should be noted that the differences truly reflected phenotypic variation rather than variations in signal intensity because cell concentrations were normalised before the FT-IR analysis.

PLSR for modelling relationships between extracellular product formation and metabolic fingerprints The aerobic growth of *E. coli* on glucose results in the production of acetate due to partial oxidation of the glucose [43]. Accumulation of acetate causes growth inhibition and affects the production of recombinant proteins [44, 45] because it acts as an uncoupler [46]. Thus, undissociated acetic acid can diffuse freely across the cell membrane and dissociates in the alkaline interior of the cell. This dissipates the trans-membrane pH gradient, causing uncoupling. In addition to this, the internal osmotic pressure increases due to the increased concentration of the acetate anion, and this has been shown to interfere with methionine biosynthesis (*inter alia*) [47]. To establish if there were any specific relationships between the production of acetate and the metabolic fingerprints of the cells, PLSR analysis was computed as described above for the prediction of extracellular acetate concentrations from the FT-IR spectra (4,000–900 cm^{-1}). The PLS regression model illustrating the calibration of the control fingerprints with acetate and the predicted values for the validation samples is shown in Fig. 4a. The number of latent variables was selected by the projection of the hold-out data at the point when the root mean square error is at its minimum (Fig. 4b). The findings of the PLSR analysis and the number of latent variables applied in each of the PLSR models are summarised in Table 2. The prediction of the acetate concentration was achieved with reasonable accuracy, suggesting a relationship between the acetate in the exometabolome and the phenotypic changes detected in the metabolic fingerprints. This calibration was based on the extracellular acetate produced by the cells, and the FT-IR analysis was performed on the intracellular fingerprint of the washed cells. The cells were washed to ensure that adherent metabolites were removed from the intracellular samples. The PLSR loading plot (data not shown) did not indicate any bands that could be solely attributed to intracellular acetate. Hence, we can hypothesise that the relationship highlighted by the PLSR analysis is indicative of the effects of acetate upon the phenotype of the cells rather than the intracellular accumulation of acetate, since measurements of acetic acid in the intracellular samples were not observed (data not shown).

Biotransformations in fed-batch cultures

The starting substrate (toluene) was supplied to the fed-batch cultures of *E. coli* by the addition of either a liquid or a vapour. The growth was performed as described above, and the toluene feed started after the addition of the IPTG to induce the expression of the recombinant protein. There were considerable differences in the TCG concentration between the cultures fed with vapour or liquid toluene (Table 1, Figs. 5c and 6c). The final product concentration was greater for the vapour-fed cultures (58–

91 mM) compared with cultures provided with liquid toluene. The liquid-fed cultures also displayed large variations in the product concentration (0.28–26 mM). The production of TCG was proportional to the biomass concentration, and generally the rate of TCG formation was at its maximum when the culture dry weight exceeded 2 g L^{-1}. However, when the toluene was supplied as a liquid feed, the TCG production rate did not increase until the toluene feed rate was increased to 0.5 mL h^{-1} (Fig. 6c).

The vapour-fed biotransformations

Five fed-batch cultures were grown in the presence of toluene vapour; the growth profiles and extracellular concentrations of acetate and TCG are shown in Figs. 5a–c, respectively. The induction of the recombinant protein was delayed in two of the cultures to determine the effects on product yield and the metabolic fingerprints (Table 1). The cultures with the higher initial biomass concentrations did not show any difference in the overall biomass yield compared with cultures with a lower initial biomass. Our analysis illustrates a relationship between the production of TCG and biomass yield; the culture with the highest maximum biomass had the lowest max TCG concentration (Table 1, the fifth culture). However, there was no difference in the starting biomass of the culture to explain the poor product yield, as the starting density of the culture was comparable to the other replicates.

Cluster analysis of the metabolic fingerprints Two of the cultures were monitored frequently throughout the entire experimental period to obtain accurate kinetics of growth and product formation. The cluster analysis of the metabolic fingerprints of the cells grown in the presence of vapour toluene showed a time-dependent trend across CV 1, with two distinct trajectories observed in CV 2 for four of the repeated cultures (Fig. 5d, e, illustrated in red and green [cultures 1 and 3] vs. blue and black [cultures 2 and 4]). The major changes observed in the spectrum (Electronic Supplementary Material Figure S1B) were in vibrations attributed to protein, with a positive change occurring in the region of 1,600–1,500 cm^{-1} (amide II) and a negative change in the region of 1,750–1,650 cm^{-1} (amide I), and changes were also observed in the polysaccharide region of the spectra (1,200–900 cm^{-1}). The induced variation in the starting biomass of the cultures was reflected across CV 2, with differences observed in the score plot of the metabolic fingerprints (Fig. 5d). The shift in the trajectory of the samples of culture repeats 1–4 (Fig. 5e) correlated with the kinetics of the product formation (Fig. 5c), which were similar for the first and third (red and green), and second (blue) and fourth (black) cultures and occurred as the rate of product formation changed. This was particularly evident for the fourth replicate, in which the rate of TCG increased after 8 h (Fig. 5c) at which point the trajectory of the samples changed in the PC-CVA scores plot (Fig. 5e). The fifth repeat (illustrated in magenta) followed a completely different trend from the other four cultures, and separation of the samples was not observed until the onset of stationary phase. However, this biotransformation was atypical, with a low TCG yield and a higher maximum biomass concentration, suggesting a difference in the cellular metabolism in comparison with the other cultures.

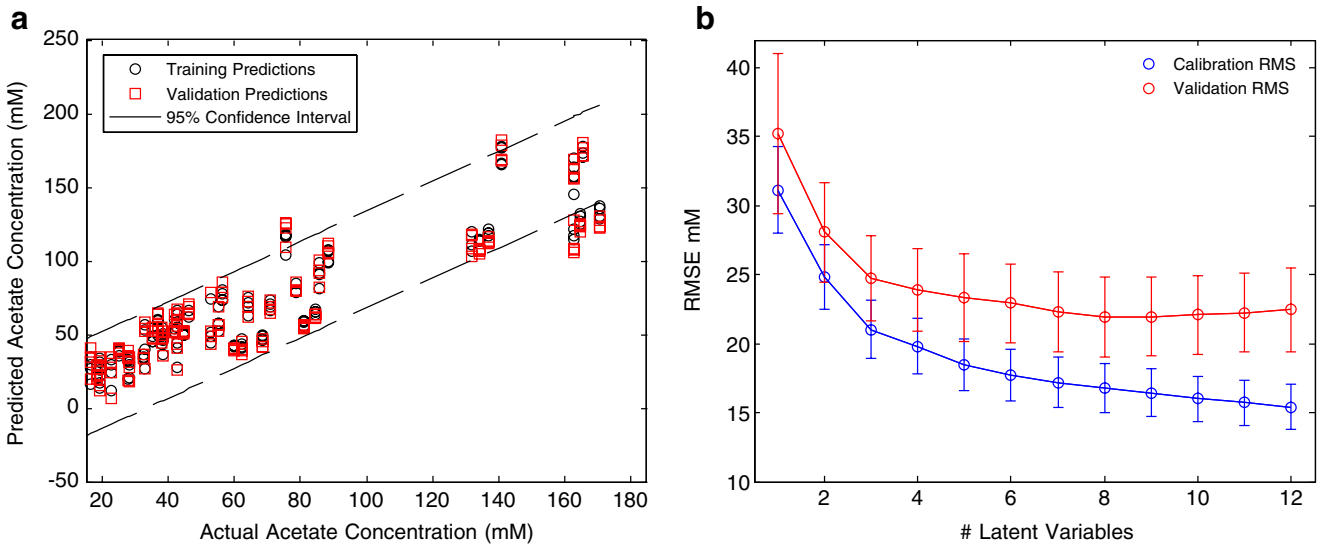

Fig. 4 PLS regression model illustrating the calibration of the control fingerprint by acetate concentration with the predictions of samples fitting the 95% confidence limits (**a**) and the RMSE error for the training and cross-validation data (**b**)

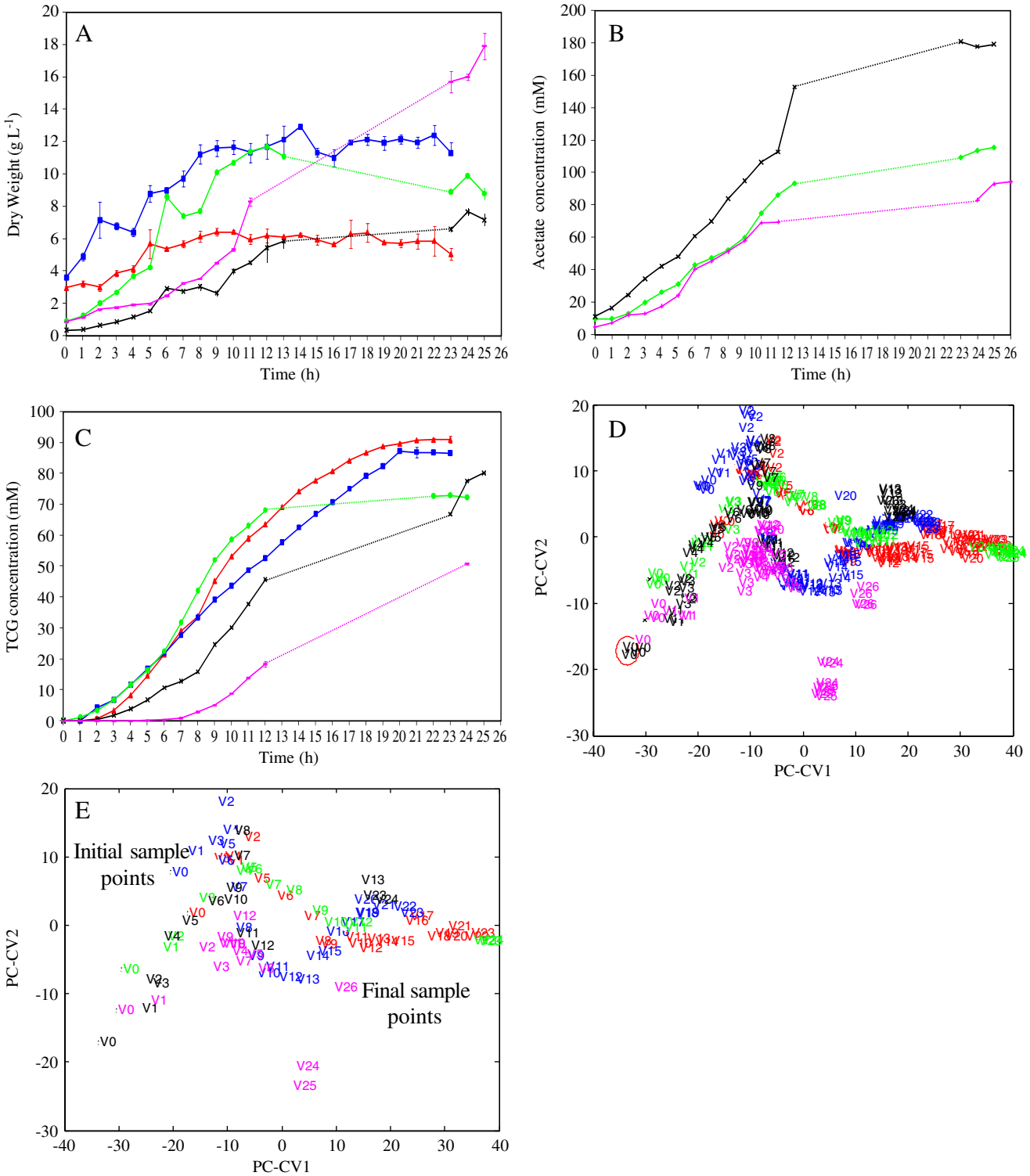

Fig. 5 The growth profile (**a**) of *E. coli* in fed-batch cultures with toluene supplied as a vapour feed, with the corresponding extracellular acetate (**b**) and TCG (**c**) concentrations. The IPTG and toluene supply was added directly after the 0-h time point. The *error bars* indicate the standard deviation of the analytical replicate data points. PC-CVA performed on the metabolic fingerprints generated by the FT-IR analysis (**d**) PCs 20 (total cumulative variance 99.86%) was used by the CVA algorithm with a priori knowledge of analytical replicates. The variance represented in the canonical variates is 9% (CV1) and 1.6% (CV2). The sampling time in hours is referred to as numbers after the letter code. For ease of visualisation, the group mean centres of each replicate are shown in **e**, whilst in **d**, all of the data points are shown and the *circle* indicates the boundary of the 95% confidence region for one particular sample. Each point would have the same area representing the 95% confidence region; however, only one circle is shown for clarity. The *colours* indicate the biological replicates (replicates 1, *red*; 2, *blue*; 3, *green*; 4, *black*; 5, *magenta*); three analytical replicates were used in the measurements of biomass, acetate and TCG concentration, and four analytical replicates were used in the FT-IR analysis

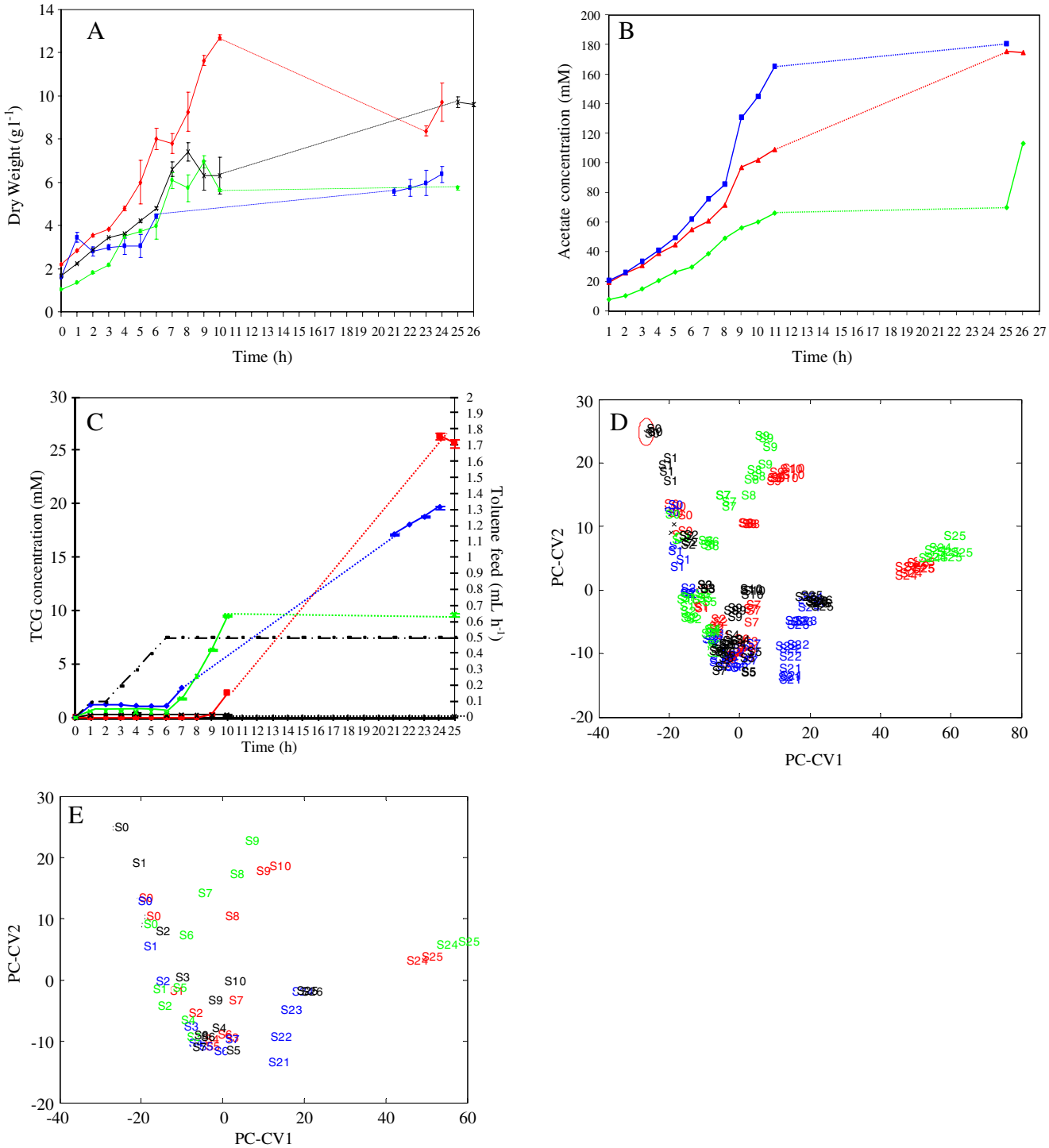

Fig. 6 The growth profile (**a**) of *E. coli* in fed-batch cultures with toluene supplied as a liquid feed, with the corresponding extracellular acetate (**b**) and TCG (**c**) concentrations. The IPTG and toluene supply was added directly after the 0-h time point. The rate of toluene feed is shown (**c**, *broken lines*). The *error bars* indicate the standard deviation of the analytical replicate data points. PC-CVA performed on the metabolic fingerprints generated by the FT-IR analysis (**d**) PCs 20 (total cumulative variance 99.92%) was used by the CVA algorithm with a priori knowledge of analytical replicates. The variance represented in the canonical variates is 22.5% (CV1) and 6.7% (CV2). The sampling time in hours is referred to as numbers after the letter code. For ease of visualisation, the group mean centres of each replicate are shown in **e**, whilst in **d**, all of the data points are shown, and the *circle* indicates the boundary of the 95% confidence region for one particular sample. Each point would have the same area representing the 95% confidence region; however, only one circle is shown for clarity. The *colours* indicate the biological replicates (replicates 1, *red*; 2, *blue*; 3, *green*; 4, *black*); three analytical replicates were used in the measurements of biomass, acetate and TCG concentration, and four analytical replicates were used in the FT-IR analysis

PLSR for modelling the relationship between extracellular products and metabolic fingerprints To investigate whether the products of the biotransformation and growth had a direct effect on the phenotype of the cells, PLSR models were computed once again. In the first instance, the TCG concentration was applied to calculate the model. The final concentration of TCG was 73.03 mM or higher in the biotransformations fed with toluene vapour. The analysis illustrated a relationship between the phenotypic analysis of the cells as detected by the FT-IR analysis and the extracellular concentration of the TCG (Table 2). This suggests that the higher concentrations of TCG observed in the vapour-fed biotransformations had an effect on the phenotype of the cells. We also modelled the relationship between the extracellular concentrations of acetate and the metabolic fingerprints. A reasonable fit was observed between the calibration variables (TCG) and metabolic fingerprints (spectra); however, the difference in the error between the calibration and validation data sets (Table 2) was larger in the acetate analysis than compared with the TCG.

The liquid-fed biotransformations

A total of four fed-batch cultures were performed with the toluene added as the liquid. The starting concentrations of biomass (Fig. 6a) were on average higher than in the vapour-fed cultures to minimise the toxic effects of adding liquid toluene. The production of TCG in the cultures fed with liquid toluene was highly variable (Table 1, Fig. 6c) and considerably lower than in the vapour-fed cells. In the early stages of the biotransformation (0–4 h), the flow rate of liquid toluene was gradually increased to 0.5 mL h^{-1} (Fig. 6c). The production of TCG did not increase until the toluene was added at its maximum rate (Fig. 6c).

Cluster analysis of the metabolic fingerprints A gradual trend was observed in the CVA space as the biotransformations progressed and the rate of the liquid toluene feed was increased (Fig. 6). The changes highlighted in the cluster analysis mirrored differences observed in the product formation of the four cultures (Figs. 6c, d). A shift in the trajectory of the samples from culture repeats 1 and 3 (illustrated in red and green) was observed at 6 and 7 h, respectively, at which point the toluene feed was at its maximum rate but the TCG concentration was low. The product yield of the second repeat was comparable to the culture repeat 1 (Fig. 6c). However, whilst a shift in the trajectory of the samples was observed in repeat 1 and also in repeat 3 in relation to the onset of TCG production, the same effect was not observed for repeat 2 (Fig. 6d, shown in blue). The product yield in the fourth culture was very low, implying either poor activity or expression of toluene dioxygenase, and whilst changes were observed in the metabolic fingerprints of this culture repeat (Fig. 6e), they were not explained by the growth or product formation data. In the score plot of the analysis (Fig. 6e), the first PC-CV indicated differences between the early and late-exposed cells. The loading plot of PC-CV1 (Electronic Supplementary Material Figure S1C) showed the most prominent change in the region of 1,700–1,500 cm^{-1} of the spectra and relates to amide vibrations within the spectra. The second PC-CV (Fig. 7) illustrated changes through the time course of the experiment, and we noted widespread changes in the spectra in particular between 1,200 and 900 cm^{-1} (Electronic Supplementary Material Figure S1C) in which we observe polysaccharide vibrations.

PLSR for modelling the relationship between extracellular products and metabolic fingerprints

The PLSR analysis was used to explore the relationship between the concentrations of TCG and acetate in the exometabolome and the metabolic fingerprints of the intracellular samples. In the model calibrated with the exometabolome concentrations of TCG, the RMSE in the validation data set was slightly higher than that of the calibration data set (Table 2). TCG is not thought to have a detrimental effect on the bacterial cells until present at higher concentrations to that observed in the liquid-fed cultures, and thus, we may not observe a strong relationship between the metabolic fingerprints and TCG concentration in the exometabolome. The difference in the RMSE of the calibration and validation data sets of the PLSR model calibrated with the extracellular acetate concentration was not noticeably different and corresponds to the findings observed in the other models calibrated with acetate (Table 2). Although the analysis suggests that the variance observed in the metabolic fingerprints of the samples could be related to the effect of acetate or TCG in the exometabolome, it is likely that other factors are influencing changes in the metabolic fingerprints of the cells.

Comparison between the control, vapour- and liquid-toluene-fed cells

In order to compare the differences in the metabolic fingerprints between the control, vapour- and liquid-fed cells directly, representative examples from each experimental condition were selected and analysed. These included repeat 1 from the control cultures (Fig. 2), repeats 1 and 2 from the vapour-fed cultures (Fig. 5) and repeat 1 from the liquid-fed cultures (Fig. 6). A single

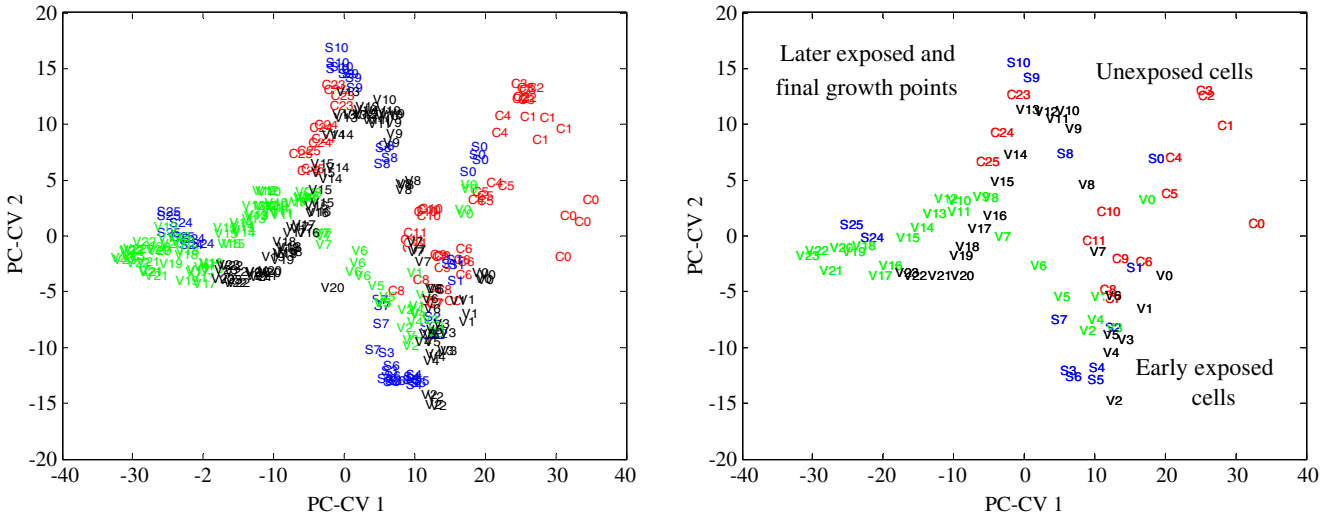

Fig. 7 PC-CVA illustrating the most reproducible examples from the three experimental conditions. PCs 20 (total cumulative variance is 99.91%) were used by the CVA algorithm with a priori knowledge of experimental replicate. The *left-hand figure* shows all the data points and the *right-hand one* the average of the group centres. The variance represented in the canonical variates is 25.97% (CV1) and 5.7% (CV2). The *colours* in the figures refer to biological replicate biotransformations, the coding used is *c* for control, *v* for vapour-fed and *s* for liquid-fed cells, and the *numbers* refer to the sampling time in hours. Four analytical replicates were included at each sample point

fermentation was selected from the control and liquid-fed cells, and two representatives were selected from the vapour biotransformations, to reflect the two trajectories observed in the earlier analysis (Fig. 5d). The PC-CVA cluster analysis (Fig. 7) illustrates two general trends. CV 1 demonstrates a general trend between early (0–7 h) and late-phase samples (8–25 h), whilst CV 2 shows a trend through the defined groups. We only illustrate CV scores 1 and 2 (as in the previous analysis); the lower CV scores (3–4) do not demonstrate clustering of the samples (data not shown). The control and vapour-fed cells both show an 'M'- or 'W'-like trajectory in CV 1 and CV 2, with a clear discrimination of the specific time points between the control and toluene-exposed cells (Fig. 7). A progressive 'stretching' of the end-point samples was observed across CV 1, in which the final control samples clustered with the mid-growth-phase vapour cells (V14–15, Fig. 7). The clustering of the toluene-exposed cells was extended ('stretched') across CV 1 in relation to the TCG production. The end-point liquid-fed samples (shown in blue, Fig. 7) had the lowest TCG yield (26.38 mM) of the three toluene-exposed biotransformations (in this analysis, the TCG production for the two vapour-fed replicates was 86.9 mM (shown in black) and 91.02 mM (illustrated in green)). The highest product concentrations in the vapour-fed cultures (TCG greater than 79.27 mM) were established after 16 h, and this coincides with the 'stretching' of the end-point samples through the CVA space (data not shown). The loading plot of the analysis (Electronic Supplementary Material Figure S1D) showed widespread changes in the spectra. In PC-CV1 in which we observed the change between the unexposed and early compared with the late-exposed cells (Fig. 7), we particularly note changes in the spectral regions of 3,000–2,800 cm^{-1} which is attributed to changes in the fatty acids of the cells and regions of polysaccharides vibrations 1,200–900 cm^{-1} (Electronic Supplementary Material Figure S1D). In PC-CV2, we also observe changes in the loading plot in the regions of the spectra which are attributed to fatty acids and polysaccharides, but we also see changes in the protein region of the spectra (1,700–1,500 cm^{-1}) and between 1,450 and 1,200 cm^{-1} in which we would expect phospholipids, polysaccharides and nucleotides to vibrate. It is therefore apparent that we observe widespread changes in the cellular components of the cell in the second PC-CV of the analysis, and in comparison with the score plot of the analysis, this corresponds with changes in the time course of the experiment (Fig. 7).

Discussion

The assessment of a 'good' or 'poor' biotransformation is principally assessed according to the quantity and rate of product formation. Offline analysis of products and growth profiling provide valuable information for the characterisation of a 'good' or 'poor' biotransformation but allow limited insight into the metabolic changes occurring at the intracellular level. To date, minimal attention has been directed towards the phenotypic analysis of the cells throughout the growth period. In this investigation, we

have employed FT-IR spectroscopy as a metabolic fingerprinting technique coupled with multivariate statistical analysis to interrogate variations in the metabolic fingerprints of *E. coli* cells during the biotransformation of toluene to toluene *cis*-glycol. The analysis illustrated specific patterns which could be attributed to variations in the growth of the cultures or product formation. The changes observed in the cultures reflect both the biological variance between the culture repeats and induced changes in the experimental protocol, which were applied to evaluate the method. Relationships were observed between the variance in the metabolic fingerprints and increasing concentrations of both the biotransformation product and acetate.

The analysis of the control cells formed a baseline for the metabolic fingerprinting of the biotransformation, with the observation of phenotypic changes in the cultures as the time course progressed. The control of the nutrient feed rates was adaptive, according to the performance of each culture, and as such variance was observed between the culture repeats. We explored the dynamics of the variation observed within and between classes with an approach previously demonstrated by us in the production of gibberellic acid by fed-batch cultures of the fungus *Gibberella fujikuroi* [18]. Changes in the actively growing cultures were reflected in the analysis; however, some of the observed changes in the analysis could not be attributed to variations in growth or acetate production, and this highlights the benefit of including an alternative approach to evaluate cultures in bioprocesses.

The production of TCG is dependent on actively growing cells and cofactor regeneration [3]. Differences in the production of TCG were observed between the vapour- and liquid-fed biotransformations, with the product yield being higher and more reproducible in the vapour-fed cultures. The initial rate in the liquid feed was low to avoid poisoning the cells, and this may have limited the TCG production in the early stages of growth. The clustering patterns observed in the vapour- and liquid-fed cultures showed differences across the time course of the experiment and between the repeated cultures, and this probably reflects the effect of both the substrate and product concentration on the cells. The PLSR analysis demonstrated strong relationships between the variance in the metabolic fingerprints and extracellular concentrations of acetate and TCG. Acetate is inhibitory in recombinant protein-producing cells [48], and whilst a relationship was noted between changes in the metabolic fingerprints and the concentrations of extracellular acetate, we did not observe a reduction in the TCG yield of cultures with a high acetate output. We suggest that the relationship highlighted by the PLSR analysis was due to metabolic changes in the cells or stress effects upon the cells since the detection of acetic acid in the intracellular samples was not observed in the FT-IR spectra. The progression of growth in any culture may be accompanied by a 'stress' effect on the cells [49], and in fed-batch growth, this may be due to nutrient limitation and accumulation of toxic by-products towards the end of the incubation period. Ageing of the culture will also affect the biotransformation process, as the growth rate decreases, and the rate of cofactor regeneration limits the activity of the dioxygenase enzyme [50]. The observed changes in the phenotype of the biotransforming cells are a combination of effects from both the substrate and product, whereas the effect on the liquid-fed cultures was predominantly related to an increase in the flow rate of toluene. The use of FT-IR as a metabolomics tool to detect the effect of chemical stresses in microbial cells was previously demonstrated [51].

The product (TCG) is inhibitory to the cells; the viability of *Pseudomonas putida*-oxidising toluene was reduced at concentrations of 79.27 mM [52]. *Pseudomonas* species are generally more tolerant to toxic substrates than *E. coli*, and changes in the phenotype of the cells were detected in relation to product (TCG) concentration (Table 2). This effect of increasing concentrations of TCG was more pronounced in the vapour-fed cells (Table 2), and this is likely the result of higher product yields. The fed-batch cultures fed with vapour toluene produce the highest product yield and would be the obvious method in the bioprocess; however, variability was still observed in the TCG yield between the repeated cultures. The approach adopted in this investigation provided clear discrimination of the 'poorest' culture repeat at an early stage in the incubation period.

Concluding remarks

We have demonstrated that metabolic fingerprinting of the cells by FT-IR analysis provides valuable information on the physiological changes in the cultures during the whole-cell biotransformations. We have performed complementary analysis to monitor growth changes and product formation in the cultures, and this has provided evidence that the FT-IR analysis reflects the observed changes in the cultures. In such a global approach where metabolic fingerprints are analysed producing a biochemical snap-shot of the cells rather than the concentration of metabolites, it is not possible to provide conclusive reasons for all of the observed variance in the data. However, this study has demonstrated the usefulness of this high-throughput method in detecting significant changes in the cells. Samples exhibiting significant changes can then be targeted for more time-consuming metabolomics approaches. We have noted considerable differences in the phenotype of the cells

fed with toluene as a liquid or vapour. The technique also has the potential to differentiate samples which will result in a poor product yield from the earliest stages of the growth period. We envisage that, with further development, this method would be a valuable tool for the online analysis of redox biotransformations to maximise product formation in bioprocesses.

Acknowledgements We would like to thank the EBS committee of the UK BBSRC for the funding of this project. We thank Rebecca Parales (University of California, Davis) for providing the plasmid pDTG601A and Chris Allen (Queen's University Belfast) for providing authentic samples of TCG as analytical standards.

References

1. Schroer K, Zelic B, Oldiges M, Lutz S (2009) Metabolomics for biotransformations: intracellular redox cofactor analysis and enzyme kinetics offer insight into whole cell processes. Biotechnol Bioeng 104:251–260
2. Cornmell RJ, Winder CL, Schuler S, Goodacre R, Stephens G (2008) Using a biphasic ionic liquid/water reaction system to improve oxygenase-catalysed biotransformation with whole cells. Green Chem 10:685–691
3. Phumathon P, Stephens GM (1999) Production of toluene *cis*-glycol using recombinant *Escherichia coli* strains in glucose-limited fed batch culture. Enzyme Microb Technol 25:810–819
4. Boyd DR, Sharma ND, Allen CCR (2001) Aromatic dioxygenases: molecular biocatalysis and applications. Curr Opin Biotechnol 12:564–573
5. Hudlicky T, Luna H, Price JD, Rulin F (1990) Microbial oxidation of chloroaromatics in the enantiodivergent synthesis of pyrrolizidine alkaloids—trihydroxyheliotridanes. J Org Chem 55:4683–4687
6. Inoue A, Horikoshi K (1989) A pseudomonas thrives in high-concentrations of toluene. Nature 338:264–266
7. Winder CL, Gordon SV, Dale J, Hewinson RG, Goodacre R (2006) Metabolic fingerprints of *Mycobacterium bovis* cluster with molecular type: implications for genotype–phenotype links. Micro-SGM 152:2757–2765
8. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. Trend Biotech 22:245–252
9. Ellis DI, Goodacre R (2006) Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. Analyst 131:875–885
10. Johnson HE, Broadhurst D, Kell DB, Theodorou MK, Merry RJ, Griffith GW (2004) High-throughput metabolic fingerprinting of legume silage fermentations via Fourier transform infrared spectroscopy and chemometrics. Appl Environ Microbiol 70: 1583–1592
11. Winder CL, Carr E, Goodacre R, Seviour R (2004) The rapid identification of *Acinetobacter* species using Fourier transform infrared spectroscopy. J Appl Microbiol 96:328–339
12. Timmins EM, Howell SA, Alsberg BK, Noble WC, Goodacre R (1998) Rapid differentiation of closely related *Candida* species and strains by pyrolysis mass spectrometry and Fourier transform-infrared spectroscopy. J Clin Microbiol 36:367–374
13. Naumann D, Helm D, Labischinski H (1991) Microbiological characterizations by FT-IR spectroscopy. Nature 351:81–82
14. Maquelin K, Kirschner C, Choo-Smith LP, Ngo-Thi NA, van Vreeswijk T, Stammler M, Endtz HP, Bruining HA, Naumann D, Puppels GJ (2003) Prospective study of the performance of vibrational spectroscopies for rapid identification of bacterial and fungal pathogens recovered from blood cultures. J Clin Microbiol 41:324–329
15. Kaderbhai NN, Broadhurst DI, Ellis DI, Goodacre R, Kell DB (2003) Functional genomics via metabolic footprinting: monitoring metabolite secretion by *Escherichia coli* tryptophan metabolism mutants using FT-IR and direct injection electrospray mass spectrometry. Comp Func Genom 4:376–391
16. Wharfe ES, Winder CL, Jarvis RM, Goodacre R (2010) Monitoring the effects of chiral pharmaceuticals on aquatic microorganisms by metabolic fingerprinting. Appl Environ Microbiol 76:2075–2085
17. McGovern AC, Ernill R, Kara BV, Kell DB, Goodacre R (1999) Rapid analysis of the expression of heterologous proteins in *Escherichia coli* using pyrolysis mass spectrometry and Fourier transform infrared spectroscopy with chemometrics: application to alpha 2-interferon production. J Biotechnol 72:157–167
18. McGovern AC, Broadhurst D, Taylor J, Kaderbhai N, Winson MK, Small DA, Rowland JJ, Kell DB, Goodacre R (2002) Monitoring of complex industrial bioprocesses for metabolite concentrations using modern spectroscopies and machine learning: application to gibberellic acid production. Biotechnol Bioeng 78:527–538
19. Franco VG, Perin JC, Mantovani VE, Goicoeche HC (2006) Monitoring substrate and products in a bioprocess with FTIR spectroscopy coupled to artificial neural networks enhanced with a genetic-algorithm-based method for wavelength selection. Talanta 68:1005–1012
20. Schuster KC, Goodacre R, Gapes JR, Young M (2001) Degeneration of solventogenic *Clostridium* strains monitored by Fourier transform infrared spectroscopy of bacterial cells. J Indus Microb Biotech 27:314–321
21. Schultz C, Naumann D (1991) In vivo study of the state of order of the membranes of gram-negative bacteria by Fourier-transform infrared-spectroscopy (Ft-Ir). FEBS Lett 294:43–46
22. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277:1453
23. Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol 166:557–580
24. Zylstra GJ, Gibson DT (1989) Toluene degradation by *Pseudomonas putida* F1—nucleotide-sequence of the Todc1c2bade genes and their expression in *Escherichia coli*. J Biol Chem 264:14940–14946
25. Stephens GM, Dalton H (1987) The effect of lipophilic weak acids on the segregational stability of Tol plasmids in *Pseudomonas putida*. J Gen Microbiol 133:1891–1899
26. Wahbi LP, Gokhale D, Minter S, Stephens GM (1996) Construction and use of recombinant *Escherichia coli* strains for the synthesis of toluene *cis*-glycol. Enzyme Microb Technol 19:297–306
27. Roscini L, Corte L, Antonielli L, Rellini P, Fatichenti F, Cardinali G (2010) Influence of cell geometry and number of replicas in the reproducibility of whole cell FTIR analysis. Analyst 135:2099–2105
28. Harrigan GG, LaPlante RH, Cosma GN, Cockerell G, Goodacre R, Maddox JF, Luyendyk JP, Ganey PE, Roth RA (2004) Applications of high-throughput Fourier-transform infrared spectroscopy in toxicology studies: contribution to a study on the development of an animal model for idiosyncratic toxicity. Toxicol Lett 146:197–205
29. Goodacre R, Timmins EM, Burton R, Kaderbhai N, Woodward AM, Kell DB, Rooney PJ (1998) Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism

fingerprinting and artificial neural networks. Microb-SGM 144:1157–1170
30. Martens H, Naes T (1989) Multivariate calibration. Wiley, Chichester
31. Jolliffe I (1986) Principal component analysis. Springer, New York
32. MacFie HJH, Gutteridge CS, Norris JR (1978) Use of canonical variates in differentiation of bacteria by pyrolysis gas–liquid chromatography. J Gen Microbiol 104:67–74
33. Windig W, Haverkamp J, Kistemaker PG (1983) Interpretation of sets of pyrolysis mass spectra by discriminant analysis and graphical rotation. Anal Chem 55:81–88
34. Krzanowski WJ (1988) Principles of multivariate analysis: a user's perspective. Oxford University Press, New York
35. Wold S, Martens H, Wold H (1982) Lecture notes in mathematics. Springer, Heidelberg
36. Winson M, Goodacre R, Woodward A, Timmins É, Jones A, Alsberg B, Rowland J, Kell D (1997) Diffuse reflectance absorbance spectroscopy taking in chemometrics (DRASTIC). A hyperspectral FT-IR-based approach to rapid screening for metabolite overproduction. Anal Chim Acta 348:273–282
37. Jarvis RM, Goodacre R (2005) Genetic algorithm optimization for pre-processing and variable selection of spectroscopic data. Bioinform 21:860–868
38. Goodacre R, Neal MJ, Kell DB (1994) Rapid and quantitative-analysis of the pyrolysis mass-spectra of complex binary and tertiary mixtures using multivariate calibration and artificial neural networks. Anal Chem 66:1070–1085
39. Brereton RG (2006) Consequences of sample size, variable selection, and model validation and optimisation, for predicting classification ability from analytical data. TRAC 25:1103–1111
40. Davison AC, Hinkley DV (1997) Bootstrap methods and their application. Cambridge University Press, Cambridge
41. Efron B (1981) Nonparametric estimates of standard error—the jackknife, the bootstrap and other methods. Biometrika 68:589–599
42. Brereton R (2003) Chemometrics: data analysis for the laboratory and chemical plant. Wiley, Chichester
43. Vemuri GN, Altman E, Sangurdekar DP, Khodursky AB, Eiteman MA (2006) Overflow metabolism in *Escherichia coli* during steady-state growth: transcriptional regulation and effect of the redox ratio. Appl Environ Microbiol 72:3653–3661
44. Luli GW, Strohl WR (1990) Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. Appl Environ Microbiol 56:1004–1011
45. Han K, Lim HC, Hong J (1992) Acetic acid formation in *Escherichia coli* fermentations. Biotechnol Bioeng 39:663–671
46. Veit A, Polen T, Wendisch VF (2007) Global gene expression analysis of glucose overflow metabolism in *Escherichia coli* and reduction of aerobic acetate formation. App Microbiol Biotechnol 74:406–421
47. Wolfe AJ (2005) The acetate switch. Microbiol Mol Biol Rev 69:12
48. Koh BT, Nakashimada U, Pfeiffer M, Yap MGS (1992) Comparison of acetate inhibition on growth of host and recombinant *Escherichia-coli* K12 strains. Biotechnol Lett 14:1115–1118
49. Hoskisson PA, Hobbs G (2005) Continuous culture—making a comeback? Micro-SGM 151:3153–3159
50. Hack CJ, Woodley JM, Lilly MD, Liddell JM (2000) Design of a control system for biotransformation of toxic substrates: toluene hydroxylation by *Pseudomonas putida* UV4. Enzyme Microb Technol 26:530–536
51. Corte L, Rellini P, Roscini L, Fatichenti F, Cardinali G (2010) Development of a novel, FTIR (Fourier transform infrared spectroscopy) based, yeast bioassay for toxicity testing and stress response study. Anal Chim Acta 659:258–265
52. Jenkins RO, Stephens GM, Dalton H (1987) Production of toluene *cis*-glycol by *Pseudomonas putida* in glucose fed-batch culture. Biotechnol Bioeng 29:873–883